

# Exercício 10 [Manipulação Avançada]

\*Obrigatório

Endereço de e-mail \*

Seu e-mail

Nome Completo \*

Sua resposta

Instale e carregue o pacote dplyr. Em seguida use o seguinte comando para carregar os dados que irá trabalhar: `df=data.frame(Theoph)`. Note que Wt: é o peso do sujeito (kg); Dose: é a dose de teofilina administrada por via oral ao indivíduo (mg / kg); Time: é o tempo desde a administração do medicamento quando a amostra foi coletada (h); e conc: é concentração de teofilina na amostra(mg / L). Responda as questões abaixo usando exclusivamente o pacote dplyr. Note que ao submeter a resposta remova todos os espaços em branco. Também use um dos padrões a seguir: `filter(dataset,<<>>)` ou `dataset%>%filter(<<>>)`.

Qual o comando seleciona apenas a coluna Dose de df ? \*

`select(df,Dose)`

Qual o comando apresenta os dados para as doses maiores que 5 mg/kg ? \*

`df%>%filter(Dose>5)`



Qual o comando seleciona as linhas de 10-20 ? Dica: use o "slice". \*

```
df%>%slice(10:20)
```

Qual comando apresenta os dados para as doses maiores que 5 e cujo o tempo desde a administração do medicamento (Time) é maior que a média do mesmo? Use apenas um único comando. \*

```
df%>%filter(Dose>5 & Time>mean(df$Time))
```

Qual comando organizar df por peso (decrecente) ? \*

```
df%>%arrange(desc(Wt))
```

Qual comando organizar df por peso (crescente) e tempo (decrecente) ? \*

```
arrange(df,Wt,desc(Time))
```

Qual comando cria uma nova coluna chamada "tendencia" que é igual à Time-mean(Time)? \*

```
mutate(df,tendencia=Time-mean(Time))
```

Qual comando apresenta a maior concentração de teofilina ? Não use nenhum nome para a coluna resultante. \*

```
summarise(df,max(conc))
```



Para os exercícios abaixo, usaremos dois conjuntos de dados relacionados aos tempos de atraso de vôos do Bureau of Transportation Statistics dos EUA (X673598238\_T\_ONTIME\_REPORTING e L\_UNIQUE\_CARRIERS.csv\_). Para carregar os dados é importante que vocês usem os seguintes argumentos: quote="\"", sep = "," . Note que a extensão ".csv\_" está correta. O data frame do arquivo (X673598238\_T\_ONTIME\_REPORTING) possui apenas informações da companhia aérea por código. No entanto, queremos saber os nomes das companhias aéreas. Assim, faça o merge dos datasets "X673598238\_T\_ONTIME\_REPORTING" e "L\_UNIQUE\_CARRIERS.csv\_" através das colunas "OP\_UNIQUE\_CARRIER" e "Code". Em seguida, responda as questões abaixo.

Qual companhia teve o maior atraso ? \*

- ☐ Flair Airlines Ltd.
- ☐ Multi-Aero, Inc. d/b/a Air Choice One
- ☐ Swiss Air Ambulance
- ☐ Sky Trek International Airlines
- ☒ American Airlines Inc.

Qual companhia atrasa mais na média ? \*

- ☒ JetBlue Airways
- ☐ American Airlines Inc.
- ☐ Hawaiian Airlines Inc.
- ☐ Endeavor Air Inc.
- ☐ SkyWest Airlines Inc.
- ☐ Southwest Airlines Co.



Qual companhia atrasa menos na média ? \*

- ☐ JetBlue Airways
- ☐ American Airlines Inc.
- ☐ JetBlue Airwaysão 2
- ☐ Endeavor Air Inc.
- ☐ SkyWest Airlines Inc.
- ☒ Hawaiian Airlines Inc.

Qual companhia teve a maior proporção de atrasos ? \*

- ☒ Southwest Airlines Co.
- ☐ SkyWest Airlines Inc.
- ☐ Hawaiian Airlines Inc.
- ☐ American Airlines Inc.
- ☐ Endeavor Air Inc.



Você está encarregado de analisar um conjunto de dados que contém casos de tuberculose (TB) relatados entre 1995 e 2013, ordenados por país, idade e sexo. O recurso mais exclusivo desses dados é o seu sistema de codificação. As colunas de três a vinte e três codificam quatro partes separadas de informações em seus nomes de coluna: (i) As três primeiras letras de cada coluna indicam se a coluna contém casos novos ou antigos de TB. (ii) As próximas duas letras descrevem os tipos de casos que estão sendo contados. (iii) A sexta letra descreve o sexo dos pacientes com tuberculose. Os números restantes descrevem a faixa etária dos pacientes com TB. Carregue o conjunto de dados de <http://stat405.had.co.nz/data/tb.csv> como um novo dataframe chamado TB. Observe que o conjunto de dados TB é desordenado de várias maneiras. Primeiramente, mova os valores das colunas 3 até 23 para uma única coluna chamada de "Informacao". Dica use o `gather()`. Divida os conteúdos da variável "informacao" em cada sublinhado("\_"). Dica use o `separate()` e chame as colunas resultantes de "caso", "tipo" e "sexofaixa". Por fim, divida a variável que contém o sexo e a faixa etária ("sexofaixa") de modo a criar uma coluna de "sexo" e uma coluna de "faixa". De posse do data frame resultante, responda as questões abaixo.

Qual foi a quantidade de casos para a Tailândia (TH) de pessoas do sexo Masculino? \*

- ☐ 0
- ☐ NA
- ☐ 2279
- ☒ 223309
- ☐ 56685
- ☐ 655310



Qual a proporção de casos para os estados unidos (US) ? Não considerar valores NAs. \*

- ☒ 0.0050469892
- ☐ 0.0048729551
- ☐ 0.0045248869
- ☐ 0.0017403411
- ☐ 0.0036547163

Qual a quantidade de casos para a faixa etária 2534 do sexo feminino? \*

- ☒ 2054610
- ☐ 961304
- ☐ 661171
- ☐ 252879
- ☐ 691

Qual foi a quantidade de casos para a década de 2000 ? A década de 2000, também referida como anos 2000, compreende o período de tempo entre 1 de janeiro de 2000 e 31 de dezembro de 2009. \*

- ☒ 37300465
- ☐ 15173290
- ☐ 34617429
- ☐ 1432
- ☐ 615977



Nunca envie senhas pelo Formulários Google.

Este formulário foi criado em UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO. [Denunciar abuso](#)

Google Formulários

